

Quiles-Puchalt, N., Tormo-Mas, M.A., Campoy, S., Toledo-Arana, A., Monedero, V., Lasa, I., Novick, R.P., Christie, G.E., and Penades, J.R. (2013) A super-family of transcriptional activators regulates bacteriophage packaging and lysis in Gram-positive bacteria. *Nucleic Acids Research*. ISSN 0305-1048

Copyright © 2013 The Authors

<http://eprints.gla.ac.uk/81257/>

Deposited on: 20 June 2013

# A super-family of transcriptional activators regulates bacteriophage packaging and lysis in Gram-positive bacteria

Nuria Quiles-Puchalt<sup>1,2</sup>, María Ángeles Tormo-Más<sup>2,3</sup>, Susana Campoy<sup>4</sup>,  
Alejandro Toledo-Arana<sup>5</sup>, Vicente Monedero<sup>6</sup>, Íñigo Lasa<sup>5</sup>, Richard P. Novick<sup>7</sup>,  
Gail E. Christie<sup>8</sup> and José R. Penadés<sup>1,3,9,\*</sup>

<sup>1</sup>Instituto de Biomedicina de Valencia (IBV-CSIC), 46010 Valencia, Spain, <sup>2</sup>Centro de Investigación y Tecnología Animal, Instituto Valenciano de Investigaciones Agrarias (CITA-IVIA), Apdo. 187, 12.400 Segorbe, Castellón, Spain, <sup>3</sup>Departamento de Ciencias Biomédicas, Universidad CEU Cardenal Herrera, 46113 Moncada, Valencia, Spain, <sup>4</sup>Departament de Genètica i Microbiologia, Universitat Autònoma de Barcelona, 08193 Barcelona, Spain, <sup>5</sup>Instituto de Agrobiotecnología, CSIC-Universidad Pública de Navarra, 31006 Pamplona, Navarra, Spain, <sup>6</sup>Laboratorio de Bacterias Lacticas y Probioticos, Instituto de Agroquímica y Tecnología de Alimentos-CSIC, 46980 Paterna, Valencia, Spain, <sup>7</sup>Skirball Institute Program in Molecular Pathogenesis and Departments of Microbiology and Medicine, New York University Medical Center, 540 First Avenue, New York, NY 10016, USA, <sup>8</sup>Department of Microbiology and Immunology, Virginia Commonwealth University School of Medicine, Richmond, VA 23298-0678, USA and <sup>9</sup>Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, G12 8TA Glasgow, UK

Received January 22, 2013; Revised April 27, 2013; Accepted April 30, 2013

## ABSTRACT

The propagation of bacteriophages and other mobile genetic elements requires exploitation of the phage mechanisms involved in virion assembly and DNA packaging. Here, we identified and characterized four different families of phage-encoded proteins that function as activators required for transcription of the late operons (morphogenetic and lysis genes) in a large group of phages infecting Gram-positive bacteria. These regulators constitute a super-family of proteins, here named late transcriptional regulators (Ltr), which share common structural, biochemical and functional characteristics and are unique to this group of phages. They are all small basic proteins, encoded by genes present at the end of the early gene cluster in their respective phage genomes and expressed under *cl* repressor control. To control expression of the late operon, the Ltr proteins bind to a DNA repeat region situated upstream of the *terS* gene, activating its transcription. This involves the C-terminal part of the Ltr proteins, which control specificity for the DNA repeat region. Finally, we show that the Ltr proteins are the only phage-encoded proteins

required for the activation of the packaging and lysis modules. In summary, we provide evidence that phage packaging and lysis is a conserved mechanism in Siphoviridae infecting a wide variety of Gram-positive bacteria.

## INTRODUCTION

Prophages are temperate bacteriophages that are integrated into the chromosomes of their bacterial hosts. Whole-genome sequencing projects have revealed that prophage sequences are common and widespread among bacterial genomes. Prophages are not only important as genetic elements that impart bacterial genome variability but they may also confer a diverse array of phenotypic traits to their hosts (sometimes referred to as lysogenic conversion factors), including those that govern the course and pathobiology of bacterial infections [reviewed in (1)].

Within the Firmicute division of Gram-positive bacteria, temperate bacteriophages are key vectors for the horizontal transfer of virulence genes. In *Staphylococcus aureus*, bacteriophages encode and mobilize an impressive array of immune evasion functions as well as clinically relevant toxins, including Pantone-Valentine leukocidin, staphylokinase, enterotoxin A and exfoliative toxin (1). In addition to these phage-encoded

\*To whom correspondence should be addressed. Tel: +34 96 339 17 60; Fax: +34 96 369 08 00; Email: jpenades@ibv.csic.es

virulence factors, staphylococcal phages induce replication of *S. aureus* pathogenicity islands (2,3) and provide the packaging machinery used for these elements to be efficiently transferred [for review see (4)]. Several bacteriophage-encoded virulence determinants also contribute to pathogenesis in group A *Streptococcus*. Genome sequencing of serotype M1, M3 and M18 strains revealed many virulence factors that were encoded by phages or phage-like elements (5). In *Bacillus anthracis*, stable infection of strains with environmental bacteriophages conferred survival phenotypes in soil and earthworm intestinal niches (6), again confirming the important role of bacteriophages in bacterial biology.

Although a great deal is known about mechanisms controlling prophage induction and temporal control of bacteriophage gene expression, much of this work has been focussed on a relatively small number of model phages, which primarily infect *Escherichia coli* (7,8). These phages have been adopted as models with the assumption that their regulatory mechanisms are generally representative in phage biology. However, the diversity of phages and the large number of phage genes of unknown function suggest that there are uncharacterized regulators still to be discovered. A case in point was the recent identification and characterization of RinA, the regulator of the morphogenesis genes in a number of phages infecting Gram-positive bacteria. This regulator had gone unnoticed because it had no homologues among known phage or bacterial transcription factors. We demonstrated that RinA binds to a tightly regulated promoter region, situated upstream of the *terS* gene, and controls expression of the morphogenetic and lysis modules of the phage by activating their transcription. As expected, *rinA* deletion eliminated formation of functional phage particles and decreased transfer of phage and phage-dependent mobile genetic elements (9).

RinA homologues were found in phages infecting a wide variety of Gram-positive bacteria (9). Although these proteins had varying degrees of similarity, they shared several common characteristics: all were basic proteins ( $pI > 8$ ) with a size  $\sim 130$  aa. In addition, the *rinA* gene was generally located just upstream of the gene encoding the small terminase subunit gene (*terS*). Despite the relatively broad distribution of RinA, however, most of the sequenced phages infecting Gram-positive bacteria did not encode RinA homologues, suggesting the existence of additional regulators or mechanisms involved in activation of the late operons in these phages.

In this work, we explored the hypothesis that most of the phages infecting the Gram-positive Firmicutes control packaging and lysis in a manner similar to that described for RinA. We have now identified five different families of phage-encoded regulators controlling phage-mediated late gene expression. Each family shares the size and charge characteristics previously found for the RinA proteins and a structurally similar C-terminal domain, as well as the location of the gene in the phage genome (Figure 1). These activators constitute a super-family of proteins that we have designated as for late transcriptional regulators (Ltr). Representatives of each family have been

shown to control expression of phage-mediated packaging and lysis functions, which are required not only for phage growth but also for phage-mediated horizontal gene transfer.

## MATERIALS AND METHODS

### Bacterial strains and growth conditions

Bacterial strains used in these studies are listed in [Supplementary Table S1](#). Bacteria were grown at 37°C overnight on TSA agar medium, supplemented with antibiotics as appropriate. Broth cultures were grown at 37°C in TSB broth with shaking (240 r.p.m.).

For SOS-dependent prophage induction, bacteria were grown in TSB broth to  $OD_{540} = 0.3$  (for *S. aureus*) or  $OD_{540} = 0.15$  (for *Enterococcus faecalis*) and induced by adding mitomycin C (MC, 2 µg/ml). Cultures were grown at 32°C with slow shaking (80 r.p.m.). Lysis usually occurred within 3 h. Samples were removed at various time points after phage induction, and standard SDS minilysates were prepared and separated on 0.7% agarose gels, as described previously (10,11). Standard procedures were used for preparation and analysis of phage lysates, lysogens and transduction in *S. aureus*, as described previously (9).

### DNA methods

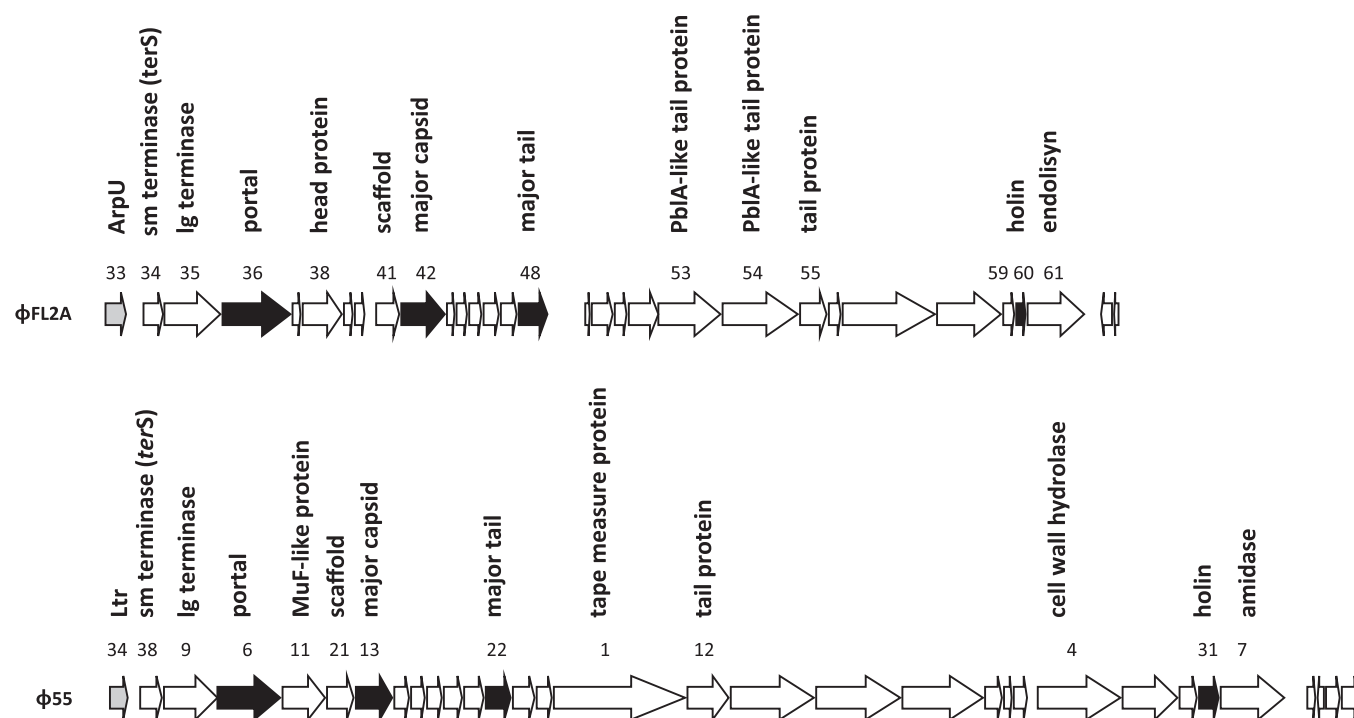
General DNA manipulations were performed by standard procedures. Labelling of the probes and DNA hybridization were performed according to the protocol supplied with the PCR-DIG DNA-labelling and chemiluminescent detection kit (Roche). To produce the strains carrying mutant prophages, allelic exchange was performed using derivatives of plasmid pMAD or pBT2 carrying the desired mutations, as described previously (11). Plasmid constructs ([Supplementary Table S2](#)) were prepared by cloning PCR products obtained from oligonucleotide primers as listed in [Supplementary Table S3](#). All constructs were sequenced by the Institute core sequencing laboratory.

### Bioinformatics analysis

Multiple sequences alignment was performed using Praline (12). The nearest neighbour tree of Ltr proteins were generated using MEGA5 (13). The hidden Markov model method SAM-T98 was used for detecting remote protein homologies (14).

### Real-time quantitative PCR

Total *S. aureus* RNA was prepared using the Fast RNA-Blue kit (Bio101) according to the manufacturer's instructions. Two micrograms of each RNA were subjected, in duplicate, to DNase I (Invitrogen) treatment for 30 min at 37°C. The enzyme was inactivated at 65°C in the presence of EDTA. To verify the absence of genomic DNA in every sample, the RNA duplicates were reverse transcribed in the presence and absence of M-MLV Reverse Transcriptase (Invitrogen). All preparations were purified using QIAquick PCR purification kit (Qiagen).



**Figure 1.** Location of the genes controlled by the Ltr proteins. Partial genetic maps of  $\Phi$ FL2A (Genbank accession number NC\_013643) and  $\Phi$ 55 (accession number AY954963), showing the location of the *ltr* genes and the genes under their control. Arrows indicate predicted ORFs, numbered as annotated. Grey arrows indicate the regulatory genes deleted in this study, whereas black arrows indicate genes for which expression was analysed by qRT-PCR.

In all, 25 ng of each reaction product was used for a real-time quantitative PCR using the iCycler machine (Biorad) and the LC-DNA Master SYBR Green I mix (Biorad). The different genes were amplified using oligonucleotides listed in [Supplementary Table S3](#). The *gyrB* transcripts that are constitutively expressed were amplified as an endogenous control. The levels of expression of the different genes were normalized with respect to *gyrB* expression. Only samples with no amplification of *gyrB* in the minus reverse transcriptase aliquot were included in the study. To monitor the specificity, the final PCR products were analysed by melting curves and electrophoresis. In each experiment, all the reactions were performed in triplicate. The relative transcriptional levels within distinct experiments were reported as previously indicated (9). The results show the average  $\pm$  SEM of at least four independent experiments.

### Design of the tiling microarray

The NA-Staph-b520729F tiling microarray was designed in collaboration with Affymetrix (Santa Clara, CA, USA). Specifically, the microarray (format 49-7875 with 11  $\mu$ m features) contains a total of 522 406 probes, divided into two parts. The first part corresponds to the tiling sub-array, which contains a total of 384 932 probes (25 mer), which are divided in eight sets. The first set corresponds to 46 310 probes recommended by Affymetrix as controls. The following seven sets are covering both strands of (i) the *S. aureus* NCTC8325 genome (2 821 347 bp covered by 363 127 probes), the *S. aureus* pathogenicity islands (ii) SaPI1 (15 229 bp covered by 1892 probes), (iii) SaPI2 (14 753 bp

covered by 1897 probes), (iv) SaPIbov1 (15 887 bp covered by 1978 probes), (v) SaPIbov2 (27 031 bp covered by 3356 probes), (vi) the staphylococcal chromosomal cassette SCC mecIV-ACMEI (54 947 bp covered by 6938 probes) and (vii) the phage 80 $\alpha$  (43 859 bp covered by 5744 probes). Each 25-mer probe was tiled each 14 nt across the whole genome, resulting in 11-nt overlaps and a 7-nt offset of the tile between strands. This design enables a 7-nt resolution for hybridization of double-stranded targets and a 14-nt resolution for strand-specific targets. The second part of the chip corresponds to the *S. aureus* gene expression sub-array, which contains 91 164 probes grouped into 3224 probe sets (each probe set includes 14 perfect match and 14 mismatch probes). In all, 3122 probe-sets are targeting unique *S. aureus* genes, whereas 102 probe-sets are controls recommended by Affymetrix.

### cDNA synthesis, fragmentation, labelling and array hybridization and scan

Before cDNA synthesis, RNA integrity from each sample was confirmed on Agilent RNA Nano LabChips (Agilent Technologies). In all, 10  $\mu$ g RNAs were reverse transcribed using SuperScript II reverse transcriptase (Invitrogen Life Technologies) and processed following the protocol of the Affymetrix GeneChip Expression Analysis Technical Manual (P/N 702232 Rev. 2) in the presence of 6 ng/ml Actinomycin D to avoid spurious second-strand cDNA synthesis during reverse transcription reaction (15). Sense RNA corresponding to *Bacillus subtilis* poly-A *lys*, *phe*, *thr*, *trp*, *dap* genes were spiked into



sample RNA as control for labelling and hybridization steps. cDNA was digested by DNase I (PIERCE) in 10× DNase I buffer (USB-Affymetrix), and the size of digestion products was analysed in the Agilent Bioanalyser 2100 using RNA Nano LabChips to ensure that the fragmentation resulted in a majority of products in the range of 50–200 bp. The fragmented cDNA were then biotinylated using terminal deoxynucleotidyl transferase (Promega) and the GeneChip DNA labelling reagent (Affymetrix) following the manufacturer's recommendations. Biotinylated cDNA (5 µg per array) were hybridized for 16 h according to the Affymetrix protocol in a total volume of 200 µl per hybridization chamber. Following incubation, the arrays were washed and stained in the Fluidics station 450 (Affymetrix) using the protocol n°FS450\_0005. Scanning of the arrays was then performed using the GeneChip scanner 3000 (Affymetrix). A first scan of the chip was carried out with gene expression sub-array parameters followed by a second scan with tiling sub-array parameters. Intensity signals of each probe cells were computed by the GeneChip operating software and stored in cell intensity files (.CEL extension) before preprocessing and analysis.

### Microarray data analysis

Data analysis of the tiling sub-array was performed using the Tiling Analysis Software from Affymetrix (<http://www.affymetrix.com>). Output bar files containing probe signal values were converted in graphic type files (.gr extension file) to be loaded and visualized using the Integrated Genome Browser version 6.2.2 from (<http://genoviz.sourceforge.net/>) (16).

### Mobility shift assays

Electrophoretic mobility shift assays (EMSAs) were performed as described before (17), using purified Ltr proteins and a DIG-labelled DNA fragment listed in Supplementary Table S3.

### Rapid amplification of cDNA 5'-ends

Amplification of the *terS* cDNA 5'-end was performed using the 5'/3' rapid amplification of cDNA 5'-ends (RACE) Kit (Roche), according to the manufacturer's protocol.

### Enzyme assays

β-Lactamase assays, using nitrocefin as substrate, were performed as described (9), using a Thermomax (Molecular Devices) microtiter plate reader. Cells were obtained in exponential phase. β-Lactamase units are defined as (V<sub>max</sub>)/OD<sub>650</sub>.

## RESULTS

### Identification of putative regulators in different phages infecting Gram-positive bacteria

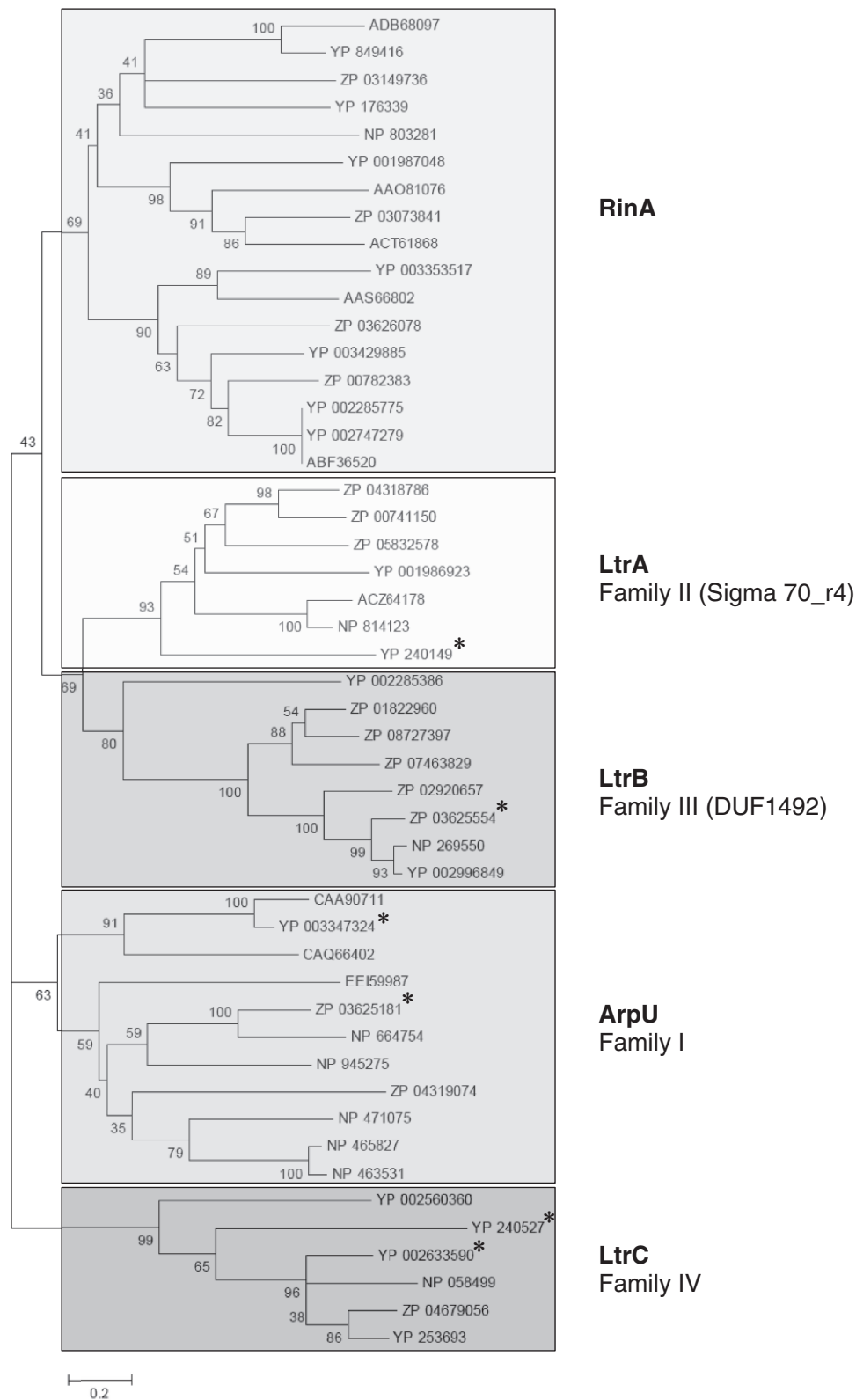
As previously indicated, RinA-like proteins are basic proteins with a length of ~130 aa, encoded by a gene located upstream of *terS*. A database search for proteins

with similar characteristics encoded at similar locations in other phage and prophage genomes revealed the existence of four additional families of proteins, distinct from RinA at the primary sequence level, encoded by phages infecting Gram-positive bacteria (Figure 2, Table 1). Although the representative members of the different families lack obvious sequence similarity (Supplementary Figure S1), using HMM-profile searches and protein folding-recognition methods, we have demonstrated that they have a conserved domain, with significant structural similarity, encompassing the C-terminal part of the proteins (Figure 3 and Supplementary Table S4). As will be demonstrated later, this domain plays an important role in the function of the proteins by recognizing the DNA repeat sequences present in the promoter region controlled by these regulators. This structural similarity, together with their common biochemical and functional characteristics, supports the idea that these proteins are members of a superfamily of related regulatory proteins. The genes for these proteins were found in phages infecting a variety of different bacterial species, including *S. aureus*, *Staphylococcus carnosus*, *E. faecalis*, *Enterococcus faecium*, *Bacillus thuringiensis*, *Bacillus cereus*, *Macroccoccus caseolyticus*, *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Streptococcus suis*, *Lactobacillus casei*, *Brevibacillus brevis*, *Listeria innocua* and *Listeria monocytogenes*. The discovery of these different families of putative regulators, including RinA, required the need for a generic term for the superfamily. Here, we propose Ltr. In addition, we also propose to designate each one of the previously uncharacterized families as Ltr followed by a capital letter.

Family I corresponded to homologues of autolysin regulatory protein (ArpU) (Table 1), and we have retained that designation. ArpU was described as a regulator of cellular muramidase-2 of *Enterococcus hirae* (18), and no additional work with this protein has been reported. Although initially described as a chromosomally encoded gene, analyses of the sequences in the database clearly demonstrate that ArpU homologues are phage-encoded (Supplementary Figure S2). Family II (LtrA) is a group of proteins with limited similarity to a portion of several sigma factors corresponding to conserved region 4 (Sigma 70<sub>r4</sub>). This region of sigma is involved in recognition of the −35 element of promoters (19), and we hypothesise that this region is involved in DNA binding and possibly specific recognition of the *terS* promoter. Family III (LtrB) corresponds to proteins carrying a conserved domain designated DUF1492. As specified in the description of this domain, this family consists of several hypothetical, highly conserved Streptococcal and related phage proteins of ~100 residues in length, the function of which is unknown. Family IV (LtrC) contained proteins homologous to the ORF34 gene product from *S. aureus* phage 55. This family has not been previously identified or characterized.

### Deletion of the genes encoding the Ltr proteins eliminated production of phage particles

To examine the functionality of the Ltr family of proteins, initial analysis of candidates from two families was carried



**Figure 2.** Ltr sequence comparison. Nearest neighbour tree of Ltr proteins generated by MEGA5 (13). Numbers indicate the bootstrap value. Asterisks indicate the regulators characterized in this study.

**Table 1.** Ltr homologs in phages from Gram-positive bacteria

Species	Accession number
Family I: ArpU homologs	
<i>E. hirae</i>	CAA90711
<i>E. faecalis</i> phage $\phi$ FL2A	YP_003347324 <sup>a</sup>
<i>E. faecium</i> TX1330	EEI59987
<i>B. cereus</i> ATCC 10876	ZP_04319074
<i>L. casei</i> BL23	CAQ66402
<i>L. monocytogenes</i> EGD-e	NP_465827
<i>L. innocua</i> Clip11262	NP_471075
<i>Listeria</i> phage A118	NP_463531
<i>Streptococcus pneumoniae</i> phage EJ-1	NP_945275
<i>S. suis</i> 89/1591	ZP_03625181 <sup>a</sup>
<i>S. pyogenes</i> MGAS315	NP_664754
Family II: Sigma 70_r4 (LtrA)	
<i>Staphylococcus</i> phage 37	YP_240149 <sup>a</sup>
<i>E. faecalis</i> phage $\phi$ FL4A	ACZ64178
<i>E. faecalis</i> V583	NP_814123
<i>E. faecium</i> C68	ZP_05832578
<i>B. cereus</i> ATCC 10876	ZP_04318786
<i>Bacillus turingensis</i> ATCC 35646	ZP_00741150
<i>L. casei</i> BL23	YP_001986923
<i>E. faecium</i> C68	ZP_05832578
Family III: DUF1492 (LtrB)	
<i>S. pneumoniae</i>	ZP_01822960
<i>S. suis</i> 89/1591	ZP_03625554 <sup>a</sup>
<i>S. pyogenes</i> NZ131	YP_002285386
<i>S. pyogenes</i> M1 GAS	NP_269550
<i>Streptococcus dysgalactiae</i>	YP_002996849
<i>Streptococcus infantarius</i>	ZP_02920657
<i>Streptococcus gallolyticus</i>	ZP_07463829
<i>Streptococcus urinalis</i>	ZP_08727397
Family IV: Homologs to ORF34 $\phi$ 55 (LtrC)	
<i>S. aureus</i> phage 55	YP_240527 <sup>a</sup>
<i>Staphylococcus</i> phage PVL	NP_058499
<i>S. carnosus</i> TM300	YP_002633590 <sup>a</sup>
<i>Staphylococcus warneri</i> L37603	ZP_04679056
<i>Staphylococcus haemolyticus</i> JCSC1435	YP_253693
<i>Macroccoccus caseolyticus</i> JCSC5402	YP_002560360
<i>Brevibacillus brevis</i> NBRC 100599	YP_002774582

<sup>a</sup>Proteins characterised in this study.

out: ArpU (Family I, from enterococcal phage  $\phi$ FL2A, YP\_003347324) and LtrC (Family IV, from *S. aureus* phage  $\phi$ 55, YP\_240527). We chose these proteins because both are encoded by transducing phages, capable of the transfer of mobile genetic elements, and they belong to the families of predicted regulators that were the most divergent from the previously characterized RinA (Figure 2). The gene for each of these putative regulatory proteins was inactivated in the corresponding  $\phi$ FL2A and  $\phi$ 55 prophages (Figure 1). The different phage mutants were induced with mitomycin C, screening lysates prepared 60 and 90 min after induction were examined by Southern blot with a phage-specific probe, and the induced cultures were assayed for phage titers.

Deletion of the genes encoding the Ltr regulators did not impair phage replication (data not shown) but eliminated the production of functional phage particles (Table 2). Furthermore, there was no lysis of the mutants after induction (data not shown), suggesting that these mutants were defective in expression of the lysis module. This defect was partially restored by

complementation; when plasmids pJP1115 (pBT2-*arpU* <sub>$\phi$ FL2A</sub>) and pJP1054 (pCN51-LtrC <sub>$\phi$ 55</sub>), carrying the *ltr* genes from  $\phi$ FL2A and  $\phi$ 55, respectively, under the control of the *P*<sub>cad</sub> promoter, were introduced into strains carrying the corresponding mutant prophage, phage titers were partially restored on mitomycin C induction (Table 2), confirming the regulatory role of these proteins in expression of the packaging and lysis functions. The complementing plasmids carry no flanking homologous DNA, and the observed phage titers for the deletion mutants depended on the presence of the complementing plasmid in the indicator strain, confirming that this was complementation and not marker rescue. We attribute the partial complementation observed in these strains to differences in timing and levels of Ltr expression from the plasmid as compared with a normal phage infection. Constitutive expression of Ltr would lead to premature expression of the phage late genes, and therefore, assembly and packaging before sufficient replication for a normal burst had occurred.

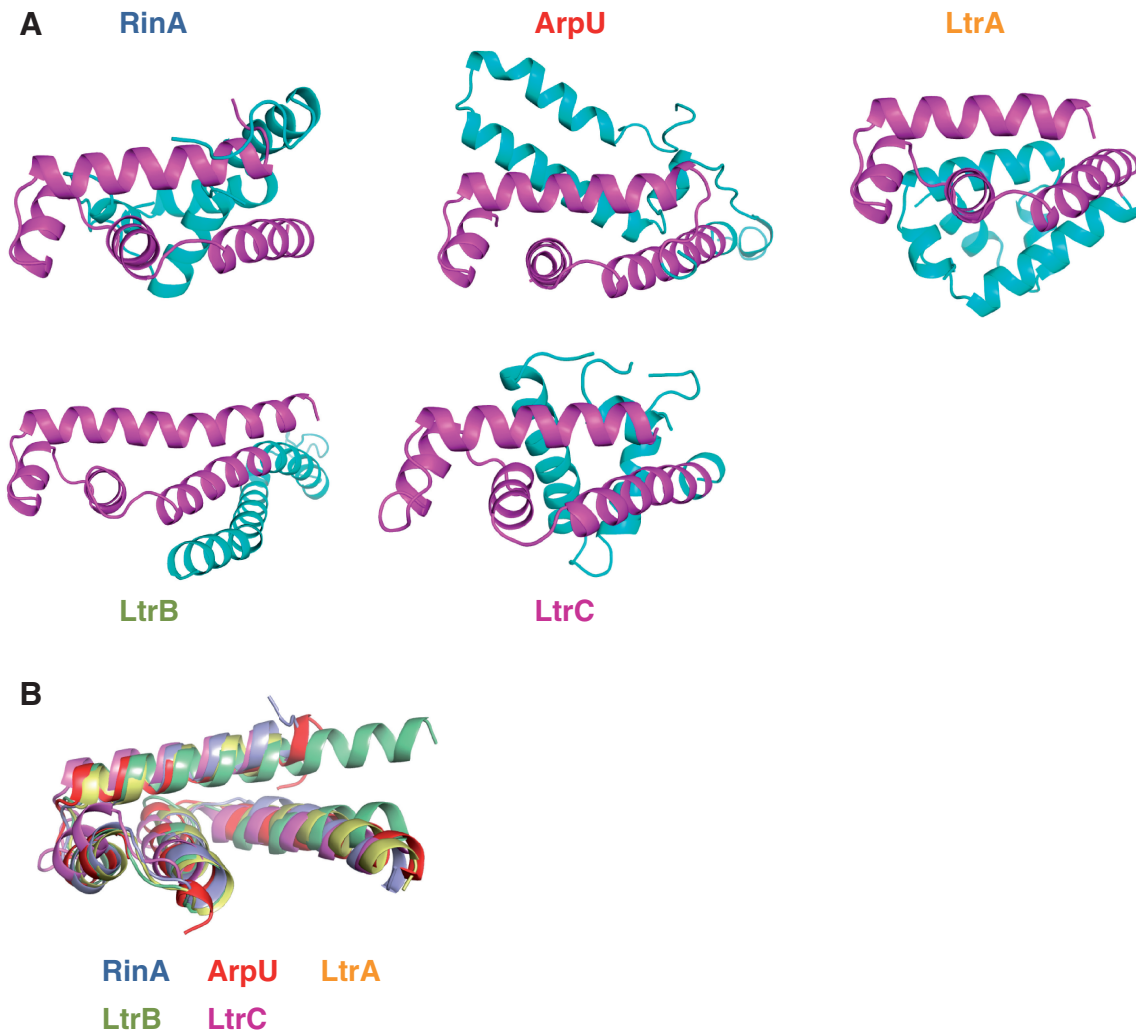
As previously reported for RinA, the phenotype of the *ltr* mutants suggested that the Ltr proteins might control phage packaging by regulating expression of the morphogenetic genes. The mRNA levels of several  $\phi$ FL2A and  $\phi$ 55 late genes (see map in Figure 1) were measured by quantitative real-time PCR at various times (0, 60 and 90 min) after MC induction, comparing expression levels in the wild-type phages and the *ltr* mutants. As a control, we also measured expression of the *ssb* and *dnaC* genes, which belong to the early transcript. The *ltr* mutation resulted in a significant decrease in the expression of genes involved in DNA packaging, capsid assembly and cell lysis, but did not affect the expression of the early genes (Figure 4). These results indicate that the Ltr proteins are transcriptional activators of the phage-encoded packaging and lysis modules.

**Identification of an Ltr-inducible promoter upstream of *terS***

As previously reported for RinA, and based on genome organization (Figure 1), we predicted the existence of at least one promoter directly controlled by the Ltr proteins, located just upstream of *terS*, the first gene in the morphogenetic gene cluster. To determine the transcription start sites and locate these putative promoters in the  $\phi$ FL2A and  $\phi$ 55 phages, 5' RACE analysis was performed with total RNA isolated after induction of the corresponding prophages. One transcriptional start site was identified for each phage, which mapped to a G located 28 ( $\phi$ 55) or 118 bp ( $\phi$ FL2A) upstream of the ATG initiation codon for *terS*.

**Ltr proteins are the only phage-encoded proteins required for *terS* expression**

To confirm the activity of the promoters upstream of the *terS* genes, and to test whether expression of the Ltr regulators was enough to induce expression of the *terS* gene in the absence of other phage-encoded proteins, we made use of plasmid pCN42, which contains a *P*<sub>cad</sub> promoter and a  $\beta$ -lactamase reporter (see Figure 5A). We cloned into this



**Figure 3.** Modelling of Ltr proteins. (A) The predicted structure of representative members of the different Ltr families is shown using a schematic (cartoon) representation coloured according to structure assignments by SAM-T98 (14). The two domains are clearly visible, with the conserved C-terminal domain coloured in magenta, and the N-terminal domain in cyan. (B) Superposition of the C-terminal domain of the different Ltr proteins. The structure of this domain is virtually identical in the different Ltr proteins. RinA (*S. aureus* phage 80 $\alpha$ ; YP\_001285353); ArpU (*E. faecalis* phage  $\phi$ FL2A; YP\_003347324); LtrA (*S. aureus* phage 37; YP\_240149); LtrB (*S. suis* 89/1591; ZP\_03625554); LtrC (*S. aureus* phage 55; YP\_240527).

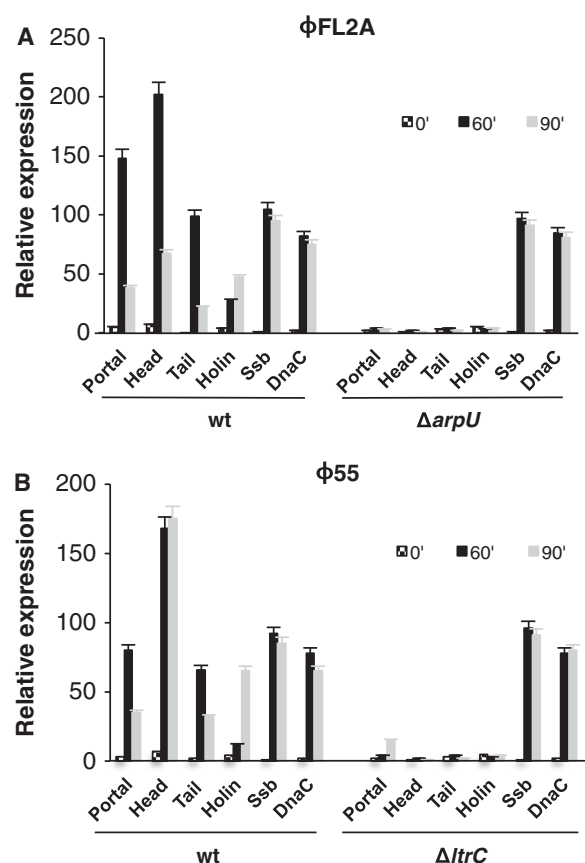
**Table 2.** Effect of mutations on phage titre

Donor strain	Phage	Plasmid	Phage titre <sup>a</sup>
<i>E. faecalis</i>			
LIV1040	$\phi$ FL2A wt		$1.3 \times 10^7$
JP8039	$\phi$ FL2A $\Delta$ arpU		<100
JP8841	$\phi$ FL2A $\Delta$ arpU	pBT2-arp U $_{\phi$ FL2A	$7 \times 10^4$
<i>S. aureus</i>			
JP2348	$\phi$ 55 wt		$1.1 \times 10^8$
JP8071	$\phi$ 55 $\Delta$ ltrC		<10
JP8172	$\phi$ 55 $\Delta$ ltrC	pCN51-ltrC $_{\phi$ 55	$9.6 \times 10^4$
JP8636	$\phi$ 55 $\Delta$ ltrC	pCN51-rinA $_{\phi$ 29	<10
JP8635	$\phi$ 55 $\Delta$ ltrC	pCN51-arp U $_{\phi$ FL2A	<10

<sup>a</sup>Pfu/ml of induced culture, using JH2-2 pCN51-arpU $_{\phi$ FL2A (for  $\phi$ FL2A) or RN4220 pCN51-ltrC $_{\phi$ 55 (for  $\phi$ 55) as recipient strains.

plasmid the *arpU* ( $\phi$ FL2A) or *ltrC* ( $\phi$ 55) coding sequences (expression of which depends on the *Pcad* promoter present in the plasmid), the *terS* promoter region and the 5' coding region of the *terS* gene corresponding to each phage, which was fused to the  $\beta$ -lactamase reporter (see Figure 5A). As controls, we generated plasmids with the same structure except that they carried mutant copies of the respective *ltr* genes (Figure 5A). These plasmids were introduced into the staphylococcal strain RN4220, and the expression of the  $\beta$ -lactamase reporter was measured. As shown in Figure 5B, both the staphylococcal  $\phi$ 55 and enterococcal  $\phi$ FL2A promoter regions were expressed in *S. aureus*, and expression of the reporter required the presence of the corresponding Ltr protein. This indicates that these regulators are the only phage-encoded proteins required for the expression of *terS*. In





**Figure 4.** Ltr proteins control expression of the morphogenetic and lysis clusters of  $\phi 55$  and  $\phi FL2A$ . Real time quantification of expression of different genes in  $\phi FL2A$  wild-type and the *arpU* mutant (A) or in  $\phi 55$  wild-type and the *ltrC* mutant (B), at different times (0, 60 or 90 min) after MC induction of the phage lytic cycle. Expression was normalized to *gyrB*.

addition, the absence of read-through expression from *Pcad* in the absence of the Ltr proteins demonstrated the existence of a transcriptional terminator in the intergenic region between the *ltr* gene and the *terS* promoter.

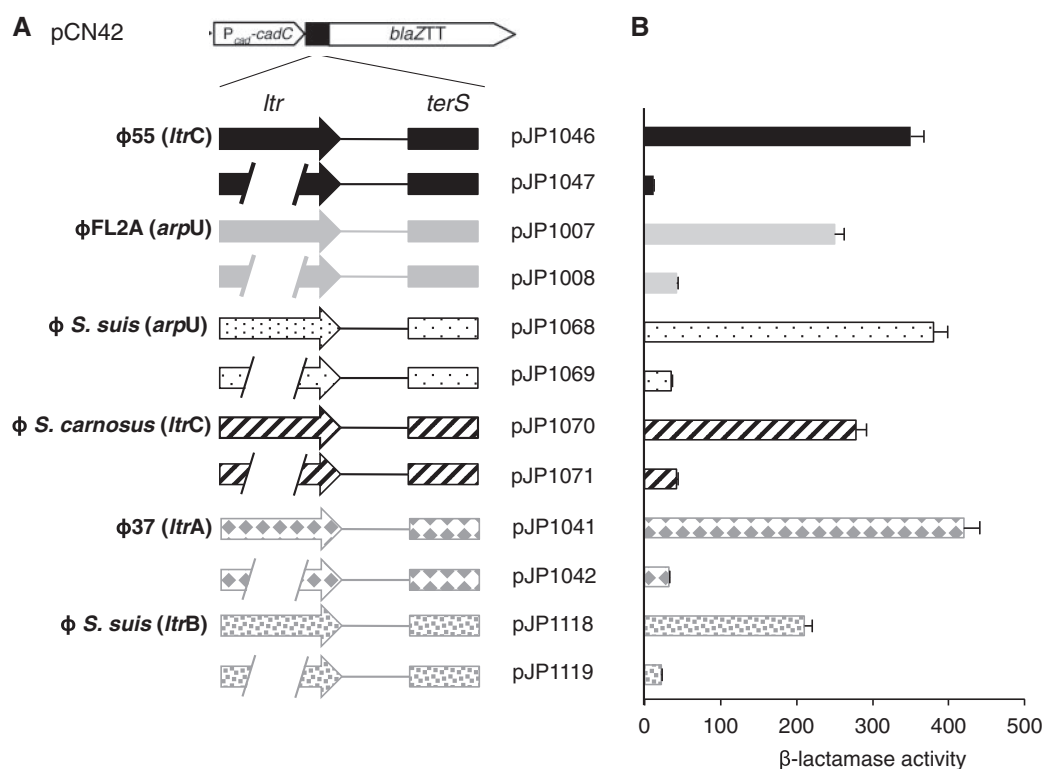
### Specificity and minimal DNA requirements for Ltr binding

The numerous genome sequences of phages encoding Ltr proteins allowed us to investigate both the specificity as well as the minimal DNA requirement that support Ltr binding. We have observed that phages expressing different Ltr proteins sometimes carry identical packaging modules. This is the case with staphylococcal phages  $\phi 55$  and  $\phi 29$  (Supplementary Figure S3). The  $\phi 29$  regulator is a RinA homologue (YP\_240603), whereas the  $\phi 55$  ORF34 protein exemplifies regulators in family IV (LtrC). We tested for possible cross-reactivity between these two regulators by determining whether RinA from phage  $\phi 29$  could activate *terS* expression from  $\phi 55$ . We cloned  $\phi 29$  *rinA* under *Pcad* control in pCN51 and introduced this plasmid into the  $\phi 55$   $\Delta ltrC$  mutant. We also complemented this mutant with the aforementioned plasmid pJP1050, a pCN51 derivative expressing ArpU. As shown in Table 2, the different regulators did not cross-

react but activated expression only of their specific *terS* genes. The intergenic region between the gene for the Ltr protein and the *terS* gene from phages  $\phi 55$  and  $\phi 29$ , which includes the terminator for their respective *ltr* genes and the *terS* promoter, is nearly identical except for the 5' region just after the *ltr* (*ltrC* or *rinA*) coding sequence (Supplementary Figure S3). These regions contain in the  $\phi 55$  and  $\phi 29$  genomes several direct repeats immediately distal to the coding sequences for their respective activator genes (Supplementary Figure S3), which we speculate are the recognition sequences for Ltr protein binding.

Initially, to determine whether the repeat regions are absolutely required for Ltr binding, we analysed this region from three different regulators (RinA $_{\phi 29}$ , LtrC $_{\phi 55}$  and ArpU $_{\phi FL2A}$ ; Figure 6A). The fragment spanning the *rinA-terS* junction from phage  $\phi 29$ , diagrammed in Supplementary Figure S3, was fused to the  $\beta$ -lactamase reporter gene present in plasmid pCN41, generating plasmid pJP1214. As a control, we generated plasmids pJP1216, pJP1217 and pJP1218, in which the nucleotides encompassing the first, the second or both repeats were changed to a stretch of As (Figure 6A). These plasmids were introduced into strains JP7904 ( $\phi 29$  lysogen), generating strains JP9585 (wt), JP9769 (first repeat mutated), JP9770 (second repeat mutated) and JP9771 (both repeats mutated), respectively. In absence of phage induction (required to provide the Ltr protein), no transcription of the reporter gene was observed in any strain (data not shown). However, prophage induction significantly increased the expression of the  $\beta$ -lactamase reporter uniquely in the strain carrying plasmid pJP1214 (wt repeats), but not in the others (Figure 6B). Identical results were obtained when we analysed, using pCN41 derivative plasmids, the role of the four repeats present in phage  $\phi 55$ . Thus, although induction of the phage  $\phi 55$  increased  $\beta$ -lactamase expression in strain carrying the plasmid pJP1221 (wt), we did not observe any  $\beta$ -lactamase expression from plasmid pJP1227, in which the DNA sequence encompassing the four repeats were changed to T (Figure 6A and B).

Finally, we observed that the phage  $\phi FL2A$ , encoding an ArpU family member, also contained three direct repeats in the same position and with the same length (7 nt) that were found for the previously characterized regulators RinA and LtrC (Figure 6 and Supplementary Figure S3). To generalize our results, we analysed the putative binding region for the ArpU protein. To do that, we made use of the aforementioned plasmid pJP1007 (see Figure 5), in which the *arpU* coding sequence under the control of the *Pcad* promoter, the *terS* promoter region, and the 5' coding region of the *terS* gene was fused to the  $\beta$ -lactamase reporter. As a control, we generated plasmid pJP1228, in which the nucleotides corresponding to the three repeats were replaced with adenines. As expected, the three repeats were absolutely required for the ArpU-dependent activation of *terS* transcription (Figure 6B). Taken together, all these results strongly suggest that the direct repeats present immediately distal to the coding sequences of their respective activator genes are the recognition sequences for Ltr protein binding.

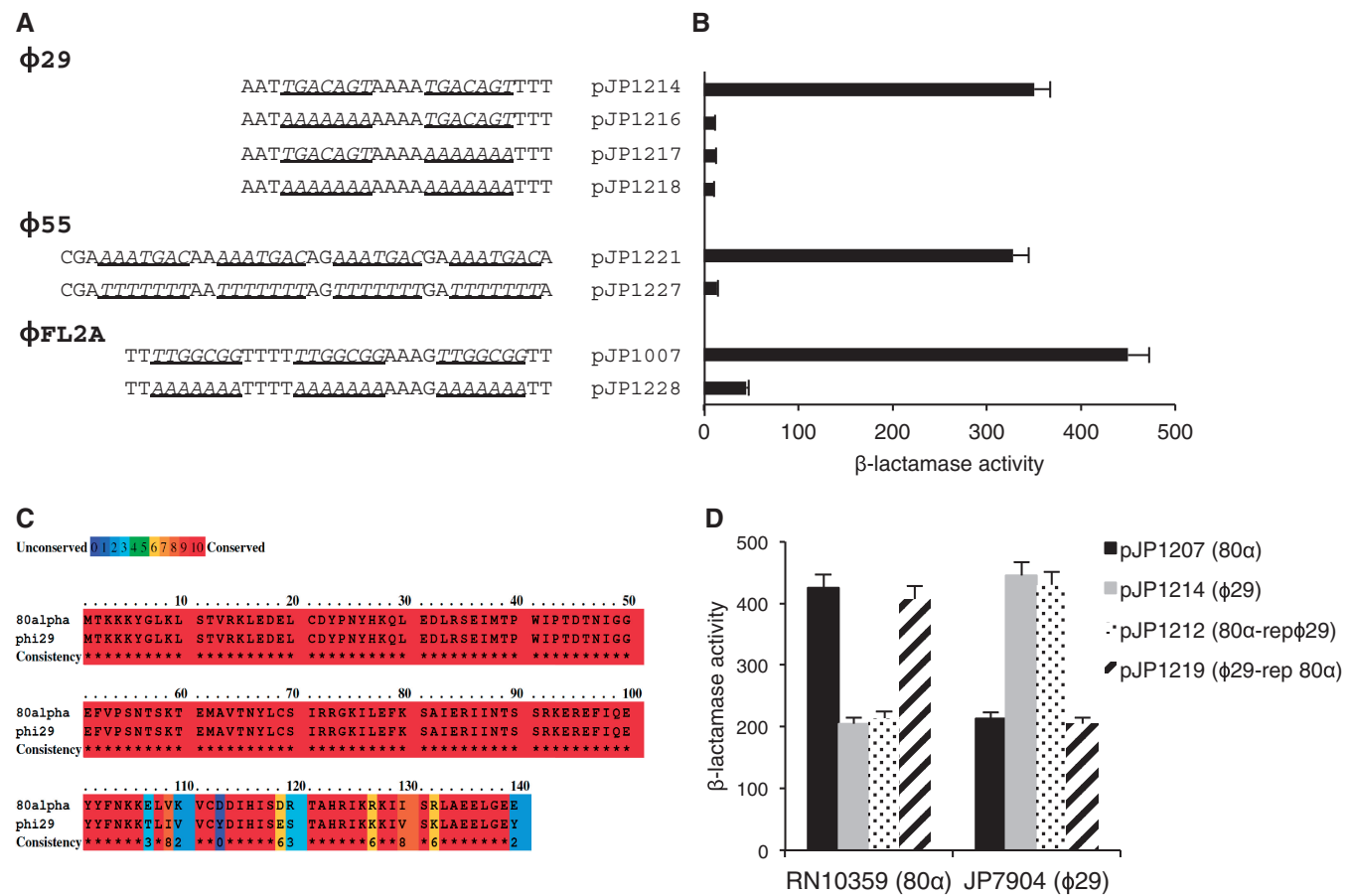


**Figure 5.** Ltr proteins control *terS* expression. (A) Schematic representation of the different *blaZ* transcriptional fusions. (B) Derivatives of strain RN4220 containing each of the indicated plasmids were assayed at mid-exponential phase for  $\beta$ -lactamase activity under standard conditions. Samples were normalized for total cell mass.  $\phi$ 55 LtrC (YP\_240527);  $\phi$ FL2A ArpU (YP\_003347324);  $\phi$  *S. suis* ArpU (ZP\_03625181);  $\phi$  *S. carnosus* LtrC (YP\_002633590);  $\phi$ 37 LtrA (YP\_240149);  $\phi$  *S. suis* LtrB (ZP\_03625554).

The results described earlier in the text do not determine whether the upstream DNA repeats are sufficient to support Ltr binding. To address this question, we initially compared activation of gene expression using plasmids pJP1207, pJP1214 and pJP1221. These pCN41 derivative plasmids contain the *terS* promoter region (including the repeats) and the 5' coding region of the *terS* gene from phages 80 $\alpha$ ,  $\phi$ 29 and  $\phi$ 55, respectively. Phages 80 $\alpha$  and  $\phi$ 29 encode a RinA protein, and as shown in [Supplementary Figure S3](#), the putative binding regions for the RinA proteins (which we have defined as a block of 59 nt just distal to *rinA*, including the direct repeats) are basically identical in these two phages. In contrast, the DNA just downstream of this block of nucleotides, including the intergenic region and the *terS* gene, is completely different. Phage  $\phi$ 55 encodes LtrC, and as shown in [Supplementary Figure S3](#), has a repeat-containing region different from that present in phages 80 $\alpha$  and  $\phi$ 29. However, as indicated previously, the cloned intergenic region from phage  $\phi$ 55 is basically identical to that present in  $\phi$ 29 and completely different from that of phage 80 $\alpha$  ([Supplementary Figure S3](#)). Plasmids pJP1207 (80 $\alpha$ ), pJP1214 ( $\phi$ 29) and pJP1221 ( $\phi$ 55) were introduced into lysogenic strains RN10359 (80 $\alpha$ ), JP7904 ( $\phi$ 29) and JP2348 ( $\phi$ 55), the resident prophages were SOS induced and the  $\beta$ -lactamase reporter was measured. As shown in [Figure 7A](#), the  $\beta$ -lactamase reporter was uniquely expressed in those strains carrying cognate pair

phage-plasmid with the DNA repeat region corresponding to the resident prophage, confirming the specificity of the Ltr proteins for their corresponding repeats. Although the RinA-expressing phages 80 $\alpha$  and  $\phi$ 29 cross-reacted, expression of the reporter gene was higher in those strains containing the specific phage and its cognate cloned DNA sequence. An explanation for this is given later in the text.

The previous results suggested that the repeat containing region is necessary and sufficient to confer Ltr binding specificity. This was further confirmed by generating chimeric structures in plasmid pCN41 in which we interchanged the repeat-containing regions of phages 80 $\alpha$  and  $\phi$ 55. As mentioned previously, although both are *S. aureus* infecting phages, the sequence of this region is completely different among these phages ([Supplementary Figure S3](#)). We generated plasmid pJP1232, which contains the repeat region from phage 80 $\alpha$  fused to the sequence present in  $\phi$ 55, and plasmid pJP1234, which contains the repeat region from  $\phi$ 55 fused to the sequence present in phage 80 $\alpha$  (see scheme in [Figure 7B](#)). A similar approach was used to generate chimeras between phages 80 $\alpha$  and  $\phi$ FL2A and  $\phi$ 55 and  $\phi$ FL2A. We generated plasmids pJP1233 and pJP1235, carrying the repeat region from phages 80 $\alpha$  and  $\phi$ 55 fused to the intergenic and *terS* DNA regions (without the repeat containing region) from the enterococcal infecting phage  $\phi$ FL2A, which encodes an ArpU family activator. The rationale was to fuse the repeat

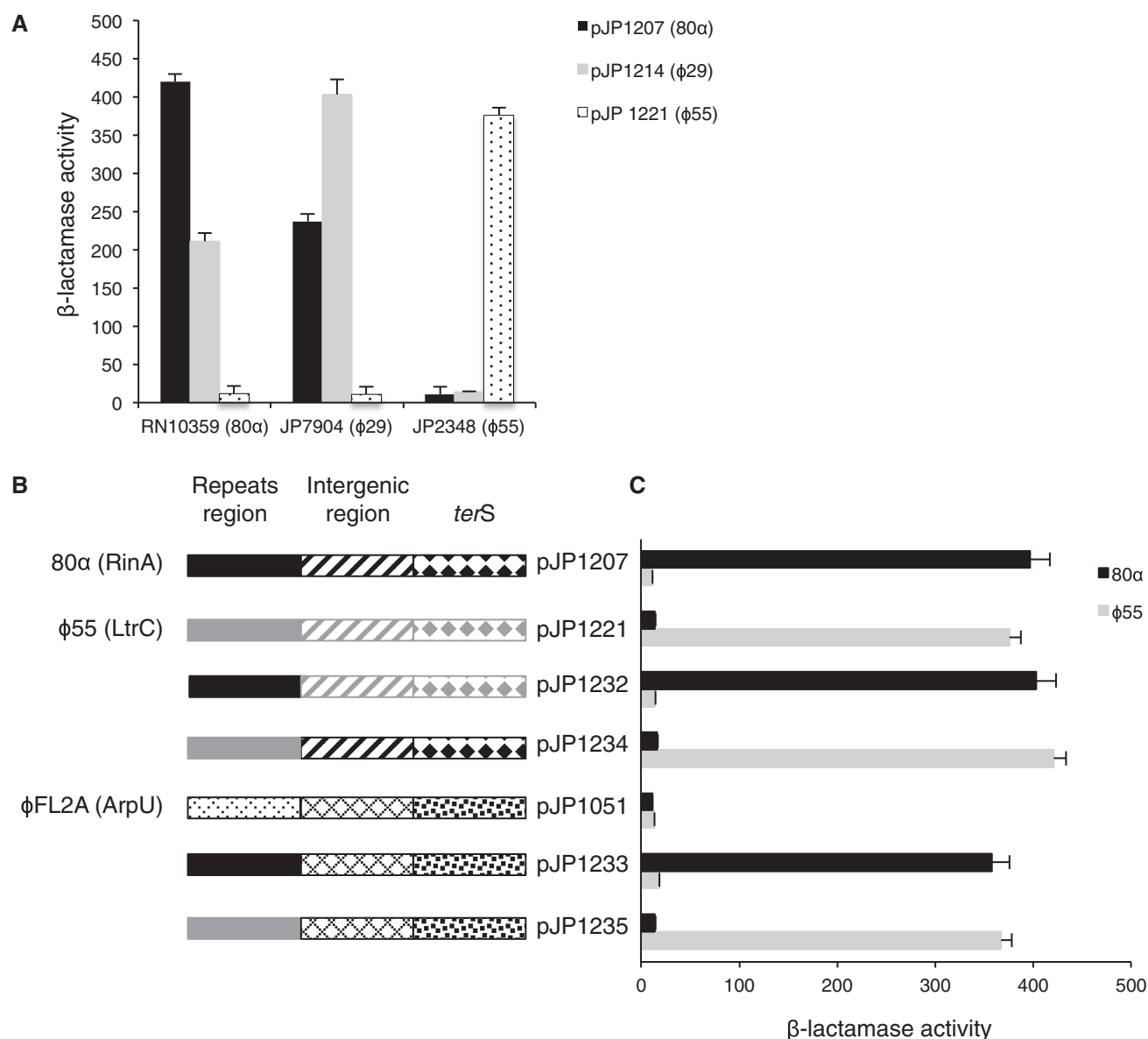


**Figure 6.** The repeat containing region is required for Ltr binding. (A) Identification of the repeat regions from phages  $\phi 29$ ,  $\phi 55$  and  $\phi FL2A$ , proposed to be involved in Ltr binding. (B) Derivatives of strains JP7904 (lysogenic for  $\phi 29$ ) and JP2348 (lysogenic for  $\phi 55$ ) containing each of the indicated plasmids were SOS induced and assayed for  $\beta$ -lactamase activity. Plasmids containing the repeats sequence from phage  $\phi FL2A$  were introduced in RN4220 and assayed at mid-exponential phase for  $\beta$ -lactamase activity under standard conditions. Samples were normalized for total cell mass. (C) Lineup of RinA protein sequences from phages 80 $\alpha$  and  $\phi 29$ , coloured according to relative sequence conservation at each position. Adapted from lineup generated by PRALINE. The scoring scheme works from 0 for the least conserved alignment position, up to 10 (asterisk) for the most conserved alignment position. (D) Derivatives of strains RN10359 (lysogenic for 80 $\alpha$ ) and JP7904 (lysogenic for  $\phi 29$ ), containing each of the indicated wt or mutated plasmids, were SOS induced and assayed for  $\beta$ -lactamase activity under standard conditions.

regions from the staphylococcal phages to DNA from other genera, which is completely different (both in sequence and in size) and is not related to that present in the staphylococcal phages (Supplementary Figure S3). These plasmids were introduced into the lysogenic strains RN10359 (80 $\alpha$ ) and JP2348 ( $\phi 55$ ), and the expression of the  $\beta$ -lactamase reporter was measured after induction of the resident prophage present in each strain. As shown in Figure 7C, phage 80 $\alpha$  induced the expression of the reporter genes uniquely from all those plasmids containing its cognate repeats. Identical results were obtained with  $\phi 55$  (Figure 7C). In summary, all these results support the hypothesis that the repeat containing region is necessary and sufficient for specific Ltr binding.

Phages  $\phi 29$  and 80 $\alpha$  encode RinA proteins that are identical except in their C-terminal regions, which contain a few substitutions (Figure 6C). As mentioned previously, when examining the repeats where these proteins bind we observed that they were identical except for a base (Supplementary Figure S3). In view of these observations, we speculated that the C-terminal

region of these proteins were different to increase their specificity for their cognate repeat sequences. To test this, we made use of the aforementioned plasmids pJP1207 and pJP1214, in which the *terS* promoter region and the 5' coding region of the *terS* gene from phages 80 $\alpha$  and  $\phi 29$ , respectively, were fused to the  $\beta$ -lactamase reporter. In addition, we generated plasmid pJP1219, which is identical to pJP1214 except that it carries repeat sequences from 80 $\alpha$ , and plasmid pJP1212, which is identical to pJP1207 except that it contains the  $\phi 29$  repeats. Each of these four plasmids was introduced into strain RN10359 (80 $\alpha$ ) and JP7904 ( $\phi 29$ ), the resident prophages were induced and the  $\beta$ -lactamase reporter was measured. As shown in Figure 6D, expression of the reporter was increased in those strains carrying the cognate pairs of phage-activator and repeats. In conclusion, our results suggest that the C-terminal region of the RinA proteins is involved specifically in recognizing the DNA binding repeats. As shown in Figure 3, the putative structure of this region was conserved in all the representative members of the Ltr superfamily analysed.



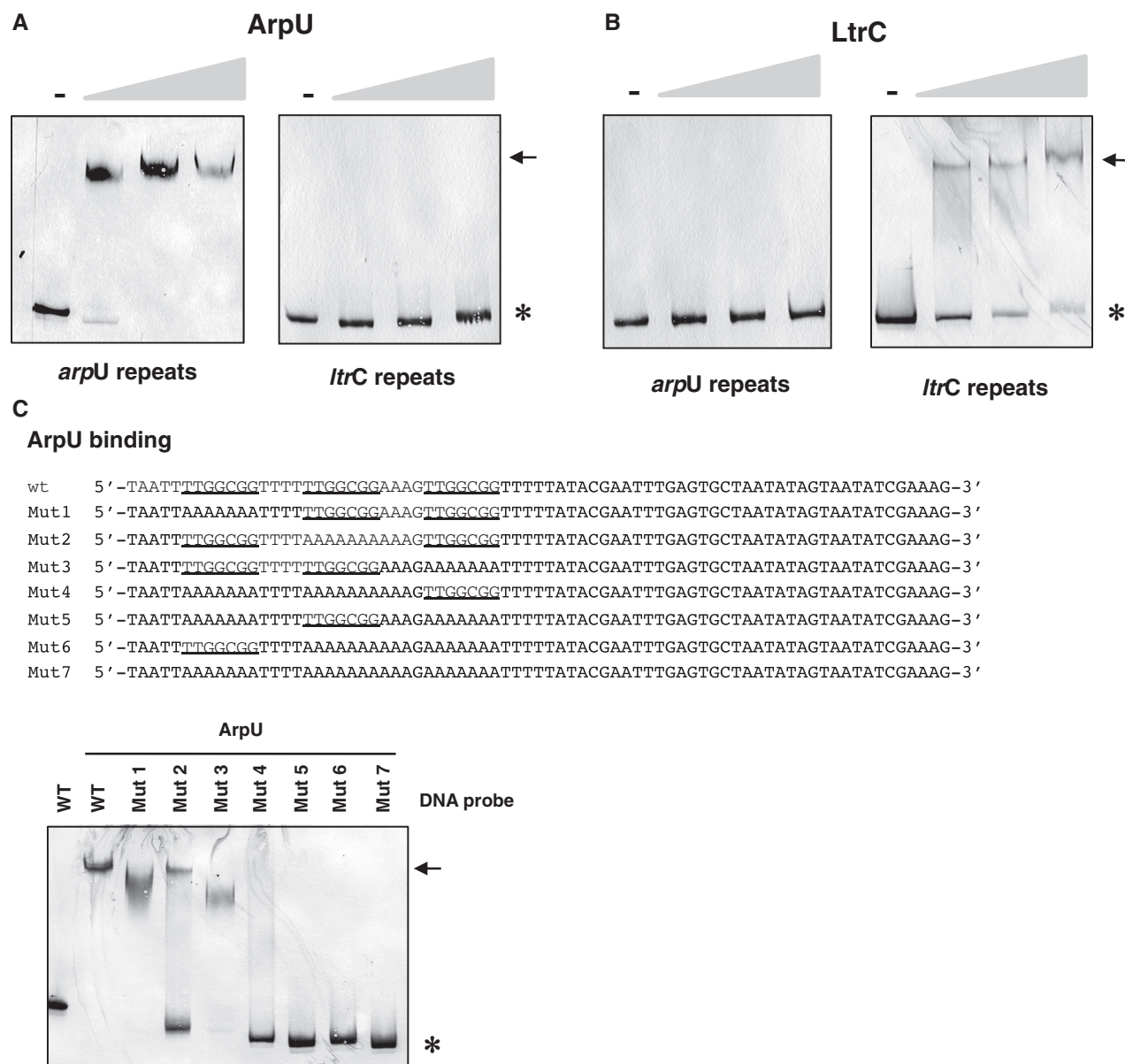
**Figure 7.** DNA requirements for Ltr binding. (A) Derivatives of strains RN10359 (lysogenic for 80 $\alpha$ ), JP7904 (lysogenic for  $\phi$ 29) and JP2348 (lysogenic for  $\phi$ 55), containing each of the indicated plasmids, were SOS induced and assayed for  $\beta$ -lactamase activity under standard conditions. (B) Scheme of the chimeric structures generated to analyse the role of the different regions contained in the *terS* promoter regions from three different phages (80 $\alpha$ ,  $\phi$ 29 and  $\phi$ 55). (C) Derivatives of strains RN10359 (lysogenic for 80 $\alpha$ ) and JP2348 (lysogenic for  $\phi$ 55) containing each of the indicated plasmids were SOS induced and assayed for  $\beta$ -lactamase activity under standard conditions.

### Ltr proteins bind to their cognate *terS* promoters

The results from the reporter plasmids suggested that the Ltr proteins bind specifically and uniquely to their cognate repeat-containing regions present at the end of the *ltr* coding sequences. To investigate this possibility further, representative His-tagged ArpU family ( $\phi$ FL2A) and LtrC family ( $\phi$ 55) proteins were expressed in *E. coli* and purified to homogeneity. Purified recombinant proteins were incubated with an end-labelled ~100-bp PCR fragment containing the entire presumptive Ltr binding region (see shaded regions in Supplementary Figure S3), and complexes were examined by electrophoretic mobility

shift assays. As shown in Figure 8, the recombinant proteins retarded mobility of their cognate *terS* promoter fragments. This binding was eliminated in the presence of an excess (500-fold) of unlabeled *terS* fragment, but not when a 500-fold excess of non-specific unlabeled DNA fragment was used as a competitor (data not shown). Moreover, although replacement of a single repeat region did not eliminate binding, the ArpU protein was unable to bind mutant derivative probes containing substitutions of two or more repeat regions (Figure 8C), further confirming that the Ltr proteins activate *terS* transcription by binding specifically to the repeat-containing region upstream of the late promoter.





**Figure 8.** Binding of the Ltr proteins to their cognate *terS* promoters. Electrophoretic mobility of DIG-labelled repeat containing region from  $\phi$ FL2A (ArpU) and  $\phi$ 55 (LtrC) was measured in the presence or absence of increasing amounts of purified ArpU (**A**) or LtrC (**B**) proteins, as indicated on the top line. (**C**) Mobility of the wild-type (wt) or mutant (mut) DIG-labelled DNA probes in the presence of purified ArpU protein. Direct repeats are underlined. Asterisk indicated DIG-labelled probe; arrows indicate the shifted complexes.

### The phage-encoded *cI* repressors control expression of the Ltr proteins

The *ltr* genes lie just downstream of the end of the phage replication module (Figure 1). The precise boundaries of the replication module have not been determined, as this region includes a number of genes of unknown function. However, it has been proposed that this module begins with *ssb*, the gene encoding a predicted single-stranded DNA-binding protein (20). It would be expected that expression of these genes, as well as *ltr*, should temporally precede expression of the morphogenetic operon and be under control of the phage *cI* repressor. In an attempt to characterize the phage transcription units, independent

northern blot analyses were performed on RNA isolated after SOS induction of phages 80 $\alpha$ ,  $\phi$ 55 and  $\phi$ FL2A. In all cases, although specific hybridization was obtained at appropriate time points, only small RNAs in the 2–4 kb size range could be detected, using probes from several different genes across this region (data not shown). This was also observed in the characterization of the morphogenetic clusters of phages 80 $\alpha$  and  $\phi$ 11 (9). This suggests that either there are multiple transcription start sites throughout this region or that the RNA is extensively processed.

To investigate further the regulation of *ltr* gene expression, we carried out transcriptional profiling using phage 80 $\alpha$  as a model. We isolated RNA before or 30 and 60 min

after SOS induction of strain RN10359, which is lysogenic for phage 80 $\alpha$ , and this RNA was analysed using tiling microarrays. The tiling microarrays were designed to analyse the genomic landscape of transcription in a strand-specific manner at 14 nt resolution (see 'Materials and Methods' section). We expected that the tiling array format would permit us to map the position of phage 80 $\alpha$  transcripts in a complementary way to the conventional molecular analysis. Figure 9A shows the tiling-array hybridization signals of phage 80 $\alpha$ , using RNA obtained at different times after induction of the phage cycle. The first observation from this experiment is that phage 80 $\alpha$  showed basal transcription of most of its genes in absence of induction. This basal transcription can be explained by the spontaneous low level of prophage induction that occurs in lysogenic cells. Second, there is a clear temporal difference between transcription of *rinA*, which is seen at 30 min post-induction, and the late (morphogenetic) operon, which does not increase until the 60 min time point. This is consistent with an essential role for *rinA* in activating late transcription. Finally, 30 min after induction of the phage lytic cycle, we observed what appears to be fairly uniform transcription on the top strand of a region encompassing the *cro* to the *rinA* genes. This region includes two predicted open reading frames (ORF11 and ORF19) that lie in the opposite orientation and would be expressed from the other strand, which is also confirmed by the tiling array. Such a distribution of mapped reads in 'overlapping operons' has been described previously in genome-wide profiling of *S. aureus* transcription (21). In that study, these overlapping reads appear to correlate with regions in which short antisense RNAs also map; therefore, it is not clear whether the apparent continuous transcription across these divergent open reading frames reflects a polycistronic mRNA that spans this region or the synthesis of an additional short transcript from the same strand that is an antisense RNA to the divergent ORF. Thus, there are two alternative interpretations of the results we obtained with 80 $\alpha$ . One possibility is that all the genes from *cro* to *rinA* belong to a single early operon, which is controlled by the phage-encoded cI repressor. An alternative interpretation is that there are multiple promoters, all co-temporally regulated, and that there are overlapping short RNAs leading to the appearance of uniform transcription across this region.

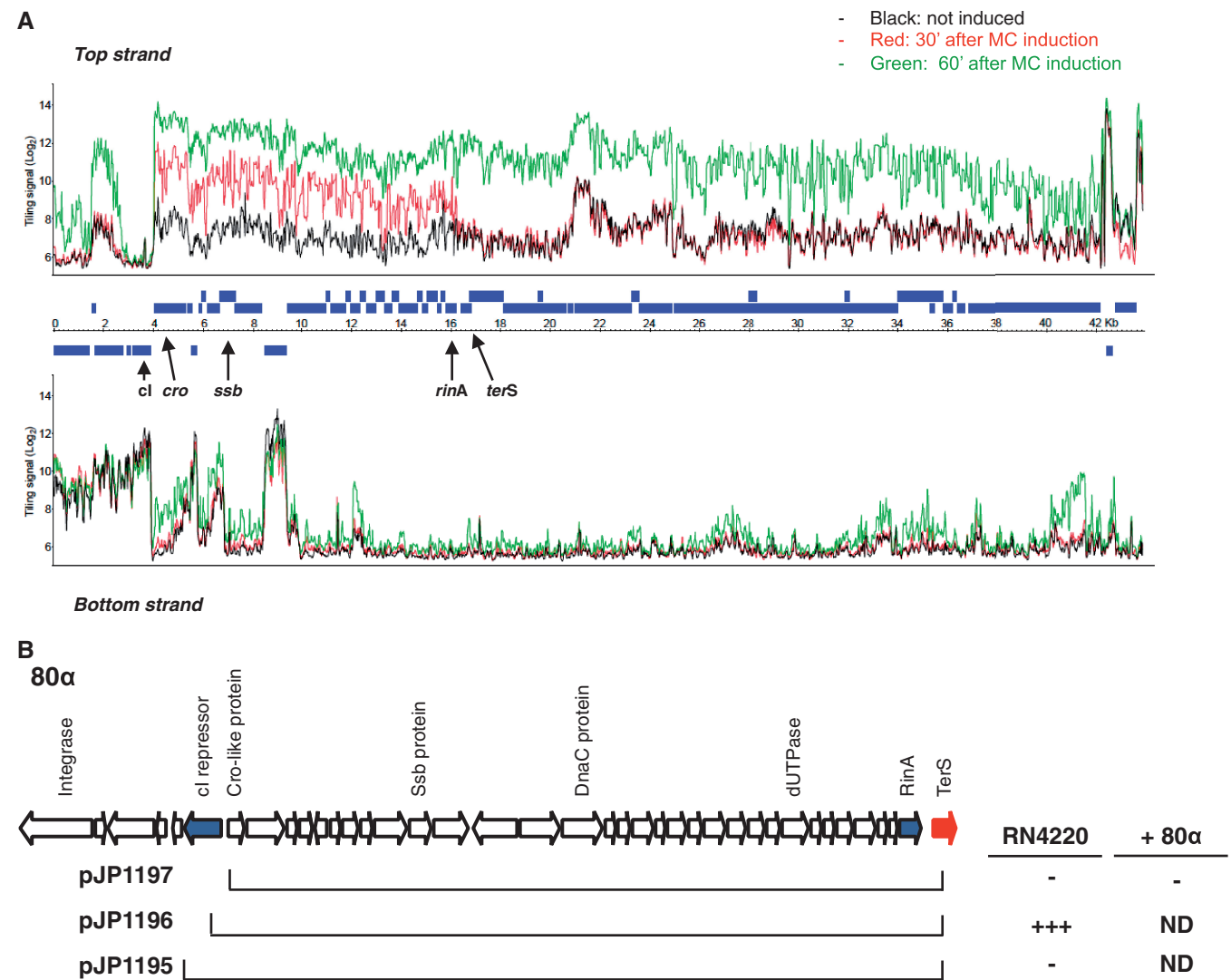
To further explore *rinA* regulation and the role of the phage repressor, we generated plasmid pJP1195, which contains the DNA region of phage 80 $\alpha$  encompassing the cI repressor to the *rinA* gene, which was fused to the  $\beta$ -lactamase reporter gene. As controls, we constructed plasmids pJP1196, which lacks the cI repressor and pJP1197, which is deleted for both the cI repressor and the cI-*cro* promoter region (see scheme in Figure 9B). Each of these plasmids was introduced into the non-lysogenic strain RN4220, and the expression of the  $\beta$ -lactamase reporter gene was measured. As expected, *rinA* transcription was observed in the strain carrying the cI-defective plasmid pJP1196, but not in strains carrying plasmids pJP1195 (encoding the cI repressor) and pJP1197 (mutant in the promoter region controlled

by the cI repressor), confirming that the cI repressor controls *rinA* expression. To determine whether this was a direct or indirect effect, we infected the strain containing plasmid pJP1197 with phage 80 $\alpha$  and measured  $\beta$ -lactamase expression. If a specific promoter other than that directly controlled by cI controlled *rinA*, 80 $\alpha$  infection should be able to restore transcription of the *rinA* gene from plasmid pJP1197 (lacking cI or the cI-*cro* promoter region). This was not the case (Figure 9B), consistent with a model in which *rinA* expression is under the direct control of the cI-*cro* promoter region and therefore part of a single polycistronic early transcript. Similar results were obtained analysing expression of *ltrC* from  $\phi$ 55 (data not shown), suggesting that the proposed mechanism controlling *ltr* transcription is widespread. It should be pointed out, however, that we cannot conclusively rule out an alternative mechanism in which cI regulates expression of a *cis*-acting factor that is required to stimulate *rinA* transcription. Regardless of the mechanism, however, these results demonstrate that *rinA* is temporally coregulated with other early genes in a cI-dependent fashion, consistent with its role in regulating late transcription.

Our previous studies of 80 $\alpha$  transcription indicated that the entire 80 $\alpha$  late gene cluster was expressed from a single *RinA*-activated promoter upstream of *terS* (9). This is consistent with what was observed in the tiling analysis of the RNA samples obtained 60 min after induction of phage 80 $\alpha$  (Figure 9A). To test whether this could be generalized to activation of the morphogenetic gene clusters of different phages by their respective Ltr proteins, we replaced the *terS* upstream region (*ltrC* and *PterS*) in strain JP2348, which is lysogenic for  $\phi$ 55, with the region encompassing *rinA* and the *PterS* from phage 80 $\alpha$ . We also replaced in strain RN10359, lysogenic for 80 $\alpha$ , the region encompassing *rinA* and *PterS* with the region that contained the *ltrC* (ORF34) and *PterS* from  $\phi$ 55. We also introduced SaPII into these strains; this SaPI is induced by both phages (3) and requires the correct expression of the entire phage-encoded late operon to be transferred (22,23). We predicted that if all late genes are co-transcribed from the promoter situated 5' of the *terS* gene, these strains will produce functional phage and SaPII particles and lyse. This was the case: the cells lysed after SOS induction and SaPII and phage particles were produced basically to the same extent as seen after induction of strains carrying wt phages (data not shown). We noted, however, that the size of the plaques was reduced in the chimeric 80 $\alpha$  phage expressing *LtrC*. A possible explanation for this is discussed later in the text.

### Ltr homologs control late gene expression in other phages from Gram-positive bacteria

In addition to the ArpU ( $\phi$ FL2A) and *LtrC* ( $\phi$ 55) regulators described earlier in the text, we tested whether other Ltr proteins, from the same family or from other families, worked similarly. Using plasmid pCN42, as previously described (Figure 5), we examined the functionality of ArpU and *LtrC* family homologues present in the



**Figure 9.** The phage-encoded *cl* repressor controls *rinA* expression. (A) Transcriptional tiling map showing expression of the early (including *rinA*) and late genes from phage 80α. Strain RN10359, lysogenic for 80α, was SOS induced, and samples were analysed at the indicated times. (B) Schematic representation of the different *blaZ* transcriptional fusions obtained to analyse the control of the *rinA* gene. Plasmids were introduced into strain RN4220 and assayed at mid-exponential phase for β-lactamase activity under standard conditions. In addition, expression of the β-lactamase reporter was also assayed after infection with phage 80α. ND; not determined.

prophages carried in *S. suis* 89/1591 (ArpU; ZP\_03625181) or *S. carnosus* TM300 (LtrC; YP\_002633590). Additionally, we analysed a member from each of the families not previously studied: the regulator present in phage 37 (YP\_240149) as an example of the LtrA (sigma70\_r4) family, and the regulator present in a *S. suis* 89/1591 prophage (ZP\_03625554) as an example of the LtrB (DUF1492) family (see scheme in Figure 5A). As a control, we generated plasmids carrying mutations in each of the different *ltr* genes. We introduced these plasmids into *S. aureus* strain RN4220, and measured β-lactamase expression. As shown in Figure 5B, all of the different Ltr proteins controlled *terS* expression, confirming that activation of transcription by members of this superfamily represents a widespread strategy used by phages infecting G+ bacteria to control phage morphogenesis and lysis.

**DISCUSSION**

Transcriptional regulation in bacteriophages infecting the low GC Gram-positive bacteria remains poorly characterized, despite the clear importance of phages in the evolution of these medically and economically important organisms. The immunity repressors of temperate phages of these bacteria have been experimentally analysed in just a few cases, including *B. subtilis* phage φ105 (24), *S. aureus* phage φ11 (25) and *Lactococcus lactis* phage TP901-1 (26). Analysis of the lysogeny modules of a collection of temperate Siphoviridae from low GC content Gram-positive bacteria revealed a common genetic organization that differs clearly from established genera of temperate Siphoviridae, as well as a novel genomic location for this module (27).



Even less work has been done on other aspects of transcriptional control, particularly in the temperate phages that play such key roles in transduction and lysogenic conversion. A promoter conserved between several lactococcal phages was shown to be regulated by a transcriptional activator from lactococcal bacteriophage  $\phi 31$  (28,29). Although this was described as a middle promoter, it is located near the phage *cos* site and lies just upstream of a gene cluster that includes a portal gene homologue, indicating that it is actually a late promoter, and analysis of the activator shows that it is a member of the RinA family. An activator of late gene transcription (Alt) has also been identified in the *L. lactis* phage TP901-1 (30). Like Ltr proteins, Alt binds to a promoter region located upstream of the *terS* gene (31), but Alt homologues appear to be much more narrowly distributed than the Ltr proteins described here; a BLAST search reveals them only in a group of related lactococcal phages. Transcriptional mapping of *S. thermophilus* phage S $\phi$ 21, based on northern blot analysis, was consistent with a role for phage-encoded regulators for middle and late transcription (32), but these were not identified.

In a previous study, we identified and characterized a family of late transcription regulators, exemplified by the RinA proteins of staphylococcal phages 80 $\alpha$  and  $\phi 11$ , which are required for expression of phage packaging and lysis functions (9). Although RinA homologues are widely distributed in phages of low GC Gram-positive bacteria, even many staphylococcal phages did not have a homologue of RinA. This prompted an *in silico* analysis of genes in the equivalent location in the genomes of these other phages, and further characterization of their function as reported here. We have now identified four additional families of transcriptional regulators that play an equivalent role in a large number of phages. All these families, including RinA, constitute a new super-family of Ltr. The *ltr* genes encode small proteins with predicted molecular weight of ~15 kDa that show no similarity to other previously characterized transcriptional regulatory proteins. All of these proteins are encoded by genes located at the end of the early gene cluster, immediately upstream of the late promoters and upstream of the late structural gene clusters. In all cases tested, these proteins appear to be activators that stimulate transcription from a promoter upstream of their respective *terS* gene, they show specificity for their cognate control regions, and they appear to be the only factor necessary for expression of the entire late gene cluster from a single promoter. It is interesting that *S. thermophilus* phage S $\phi$ 21 encodes a DUF1492 homologue (Class IV) at the equivalent position in its genome. This suggests that the multiple transcription starts reported for this phage on the basis of northern blot and primer extension analysis (32) were actually 5'-ends that arose from processing of the late transcript, as was found to be the case for  $\phi 11$  and 80 $\alpha$  (9).

An understanding of the evolution of global transcription regulators is essential for comprehending the complex networks of cellular metabolism that have developed among related organisms. Bacteriophages live under intense selection pressure, where the primary competitive

advantage is predicted to be a high replication, packaging and transfer rate (33). This situation suggests that the evolutionary selection of these organisms has optimized expression of their components to produce a maximal growth rate. Tightly orchestrated temporal control is necessary to prevent premature packaging of the bacteriophage DNA before it has been completely replicated. This temporal control is produced in most Gram-positive infecting phages by the Ltr proteins. As it is predicted that activation of the phage late promoter will occur once the Ltr protein accumulates, linking expression of the *ltr* genes to the early transcript assures that this increased expression of the *ltr* genes will be uniquely produced after efficient phage replication. Interestingly, and as discussed earlier in the text, we noted that the chimeric 80 $\alpha$  phage expressing the family IV regulator from  $\phi 55$  produced smaller plaques than wt 80 $\alpha$  phage. However, this chimeric phage produced a small number of normal plaques. Sequence analysis of three of these plaques revealed changes in the ribosome-binding site for the *terS* gene. In view of this result, we hypothesize that the different activators are optimized for different replication rates or different levels of expression of the phage-encoded proteins, and that differences in these features among the phages infecting Gram-positive bacteria could be the major force responsible for the diversification and specialization observed in the Ltr family of regulators.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–4, Supplementary Figures 1–3 and Supplementary References [34–39].

## ACKNOWLEDGEMENTS

The authors are thankful to Alberto Marina (IBV-CSIC, Valencia) for providing technical help in the modelling the Ltr proteins. J.R.P. conceived and designed the study; N.Q.-P. and M.A.T.-M. characterized the mutants and performed gene expression experiments; S.C. and V.M. performed mobility shift assay experiments; A.T.-A., M.A.T.-M. and I.L. performed the tiling array experiment; J.R.P., G.E.C., R.P.N., I.L., N.Q.-P. and M.A.T.-M. analysed the data; J.R.P. and G.E.C. wrote the manuscript; J.R.P., M.A.T.-M., G.E.C., I.L. and R.P.N. obtained funding.

## FUNDING

Funding for open access charge: Ministerio de Ciencia e Innovación (MICINN) [Consolider-Ingenio CSD2009-00006, BIO2011-30503-C02-01 and Eranet-pathogenomics PIM2010EPA-00606 to J.R.P.]; Cardenal Herrera-CEU University [Copernicus-Santander program to J.R.P.]; Instituto Nacional de Investigaciones Agrarias (INIA) [DR08-0093 to M.A.T.-M.]; National Institute of Health [R56AI081837 to G.E.C., R01AI022159-23A2 to R.P.N.].



Conflict of interest statement. None declared.

## REFERENCES

- Brüssow, H., Canchaya, C. and Hardt, W.-D. (2004) Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol. Mol. Biol. Rev.*, **68**, 560–602.
- Tormo-Más, M.Á., Donderis, J., García-Caballer, M., Alt, A., Mir-Sanchis, I., Marina, A. and Penadés, J.R. (2013) Phage dUTPases control transfer of virulence genes by a proto-oncogenic G protein-like mechanism. *Mol. Cell*, **49**, 947–958.
- Tormo-Más, M.Á., Mir, I., Shrestha, A., Tallent, S.M., Campoy, S., Lasa, I., Barbé, J., Novick, R.P., Christie, G.E. and Penadés, J.R. (2010) Moonlighting bacteriophage proteins derepress staphylococcal pathogenicity islands. *Nature*, **465**, 779–782.
- Novick, R.P., Christie, G.E. and Penadés, J.R. (2010) The phage-related chromosomal islands of Gram-positive bacteria. *Nat. Rev. Microbiol.*, **8**, 541–551.
- Banks, D.J., Porcella, S.F., Barbian, K.D., Beres, S.B., Philips, L.E., Voyich, J.M., DeLeo, F.R., Martin, J.M., Somerville, G.A. and Musser, J.M. (2004) Progress toward characterization of the group A *Streptococcus metagenome*: complete genome sequence of a macrolide-resistant serotype M6 strain. *J. Infect. Dis.*, **190**, 727–738.
- Schuch, R. and Fischetti, V.A. (2009) The secret life of the anthrax agent *Bacillus anthracis*: bacteriophage-mediated ecological adaptations. *PLoS One*, **4**, e6532.
- Court, D.L., Oppenheim, A.B. and Adhya, S.L. (2006) A new look at bacteriophage genetic networks. *J. Bacteriol.*, **189**, 298–304.
- Campbell, A. (1994) Comparative molecular biology of lambdoid phages. *Annu. Rev. Microbiol.*, **48**, 193–222.
- Ferrer, M.D., Quiles-Puchalt, N., Harwich, M.D., Tormo-Mas, M.A., Campoy, S., Barbe, J., Lasa, I., Novick, R.P., Christie, G.E. and Penadés, J.R. (2011) RinA controls phage-mediated packaging and transfer of virulence genes in Gram-positive bacteria. *Nucleic Acids Res.*, **39**, 5866–5878.
- Úbeda, C., Maiques, E., Knecht, E., Lasa, I., Novick, R.P. and Penadés, J.R. (2005) Antibiotic-induced SOS response promotes horizontal dissemination of pathogenicity island-encoded virulence factors in staphylococci. *Mol. Microbiol.*, **56**, 836–844.
- Úbeda, C., Maiques, E., Barry, P., Matthews, A., Tormo, M.Á., Lasa, I., Novick, R.P. and Penadés, J.R. (2008) SaPI mutations affecting replication and transfer and enabling autonomous replication in the absence of helper phage. *Mol. Microbiol.*, **67**, 493–503.
- Simossis, V.A. and Heringa, J. (2005) PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res.*, **33**, W289–W294.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.*, **28**, 2731–2739.
- Karplus, K., Barrett, C. and Hughey, R. (1998) Hidden Markov models for detecting remote protein homologies. *Bioinformatics*, **14**, 846–856.
- Perocchi, F., Xu, Z., Clauder-Münster, S. and Steinmetz, L.M. (2007) Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic Acids Res.*, **35**, e128.
- Nicol, J.W., Helt, G.A., Blanchard, S.G., Raja, A. and Loraine, A.E. (2009) The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics*, **25**, 2730–2731.
- Úbeda, C., Maiques, E., Tormo, M., Campoy, S., Lasa, I., Barbé, J., Novick, R.P. and Penadés, J.R. (2007) SaPI operon I is required for SaPI packaging and is controlled by LexA. *Mol. Microbiol.*, **65**, 41–50.
- Lleó, M.M., Fontana, R. and Solioz, M. (1995) Identification of a gene (*arpU*) controlling muramidase-2 export in *Enterococcus hirae*. *J. Bacteriol.*, **177**, 5912–5917.
- Dombroski, A.J., Walter, W.A., Record, M.T., Siegle, D.A. and Gross, C.A. (1992) Polypeptides containing highly conserved regions of transcription initiation factor sigma 70 exhibit specificity of binding to promoter DNA. *Cell*, **70**, 501–512.
- Christie, G.E., Matthews, A.M., King, D.G., Lane, K.D., Olivarez, N.P., Tallent, S.M., Gill, S.R. and Novick, R.P. (2010) The complete genomes of *Staphylococcus aureus* bacteriophages 80 and 80α—Implications for the specificity of SaPI mobilization. *Virology*, **407**, 381–390.
- Lasa, I., Toledo-Arana, A., Dobin, A., Villanueva, M., de Los Mozos, I.R., Vergara-Irigaray, M., Segura, V., Fagegaltier, D., Penadés, J.R., Valle, J. et al. (2011) Genome-wide antisense transcription drives mRNA processing in bacteria. *Proc. Natl Acad. Sci. USA*, **108**, 20172–20177.
- Tormo, M.A., Ferrer, M.D., Maiques, E., Ubeda, C., Selva, L., Lasa, I., Calvete, J.J., Novick, R.P. and Penadés, J.R. (2008) *Staphylococcus aureus* pathogenicity island DNA is packaged in particles composed of phage proteins. *J. Bacteriol.*, **190**, 2434–2440.
- Tallent, S.M., Langston, T.B., Moran, R.G. and Christie, G.E. (2007) Transducing particles of *Staphylococcus aureus* pathogenicity island SaPII are comprised of helper phage-encoded proteins. *J. Bacteriol.*, **189**, 7520–7524.
- Dhaese, P., Seurinck, J., De Smet, B. and Van Montagu, M. (1985) Nucleotide sequence and mutational analysis of an immunity repressor gene from *Bacillus subtilis* temperate phage phi 105. *Nucleic Acids Res.*, **13**, 5441–5455.
- Das, M., Ganguly, T., Chattoraj, P., Chanda, P.K., Bandhu, A., Lee, C.Y. and Sau, S. (2007) Purification and characterization of repressor of temperate *S. aureus* phage phi11. *J. Biochem. Mol. Biol.*, **40**, 740–748.
- Pedersen, M., Ligowska, M. and Hammer, K. (2010) Characterization of the CI repressor protein encoded by the temperate lactococcal phage TP901-1. *J. Bacteriol.*, **192**, 2102–2110.
- Lucchini, S., Desiere, F. and Brüssow, H. (1999) Similarly organized lysogeny modules in temperate Siphoviridae from low GC content gram-positive bacteria. *Virology*, **263**, 427–435.
- Walker, S.A. and Klaenhammer, T.R. (1998) Molecular characterization of a phage-inducible middle promoter and its transcriptional activator from the lactococcal bacteriophage phi31. *J. Bacteriol.*, **180**, 921–931.
- Walker, S.A., Dombroski, C.S. and Klaenhammer, T.R. (1998) Common elements regulating gene expression in temperate and lytic bacteriophages of *Lactococcus* species. *Appl. Environ. Microbiol.*, **64**, 1147–1152.
- Brøndsted, L., Pedersen, M. and Hammer, K. (2001) An activator of transcription regulates phage TP901-1 late gene expression. *Appl. Environ. Microbiol.*, **67**, 5626–5633.
- Pedersen, M., Kilstrup, M. and Hammer, K. (2006) Identification of DNA-binding sites for the activator involved in late transcription of the temperate lactococcal phage TP901-1. *Virology*, **345**, 446–456.
- Ventura, M., Bruttin, A., Canchaya, C. and Brüssow, H. (2002) Transcription analysis of *Streptococcus thermophilus* phages in the lysogenic state. *Virology*, **302**, 21–32.
- Rohwer, F. and Thurber, R.V. (2009) Viruses manipulate the marine environment. *Nature*, **459**, 207–212.
- Kreiswirth, B.N., Löfdahl, S., Betley, M.J., O'Reilly, M., Schlievert, P.M., Bergdoll, M.S. and Novick, R.P. (1983) The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. *Nature*, **305**, 709–712.
- Yasmin, A., Kenny, J.G., Shankar, J., Darby, A.C., Hall, N., Edwards, C. and Horsburgh, M.J. (2010) Comparative genomics and transduction potential of *Enterococcus faecalis* temperate bacteriophages. *J. Bacteriol.*, **192**, 1122–1130.
- Úbeda, C., Barry, P., Penadés, J.R. and Novick, R.P. (2007) A pathogenicity island replicon in *Staphylococcus aureus* replicates as an unstable plasmid. *Proc. Natl Acad. Sci. USA*, **104**, 14182–14188.
- Arnaud, M., Chastanet, A. and Débarbouillé, M. (2004) New vector for efficient allelic replacement in naturally nontransformable, low-GC-content, gram-positive bacteria. *Appl. Environ. Microbiol.*, **70**, 6887–6891.
- Charpentier, E., Anton, A.I., Barry, P., Alfonso, B., Fang, Y. and Novick, R.P. (2004) Novel cassette-based shuttle vector system for gram-positive bacteria. *Appl. Environ. Microbiol.*, **70**, 6076–6085.
- Söding, J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics*, **21**, 951–960.